MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, **2015/2016**

HPB2021 – BIOINFORMATICS PROGRAMMING II

(All sections / Groups)

02 March 2016
02:30-04:30 PM
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 6 pages including cover page.
2. Attempt **Four** out of **Five** questions. All questions carry equal marks (10), and the distribution of the marks for each question is given.
3. Please write all your answers in the Answer Booklet provided.

QUESTION 1 [10 marks]

a) Write Python code to calculate the alignment scores for the following 2 cases.

i. Global alignment: match = +1, mismatch = 0, gap = -1

S1 = "~~TCTGTACGCGATCATGT"
S2 = "TAGC~GTCCGATAT~A~~~"

ii. Local alignment: match = +1, mismatch = -1, gap = -2

S1 = "AGATAGAAACTGATATATA"
S2 = "AGA~A~A~ACAGAG~T~~~"

[5 marks]

b) The following two matrices are given:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad AT = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

i. Write Python code to give matrix A as a List.

[1 mark]

ii. Write Python function *matrix2D_Add(x, y)* to add 2 matrices like A and AT. Print the resulting matrix.

[4 marks]

Continued...

QUESTION 2 [10 marks]

a) Define a function "dig(-)" that receives a sequence and one restriction pattern and returns the positions where the enzyme cuts its pattern (Hint: the location where the restriction site starts in a sequence). You can return a list containing name of the enzyme and the position it occurs in a dna sequence.
e.g. EcoRI=['gaattc', 1,154,...] [4 marks]

b) Write a function which will calculate the GC percentage of a DNA sequence. If the name of the function is given as **gc**, it returns the following at the python interpreter: [3 marks]

e.g.>>> gc('atgtaatgatat')
16.666666666666667

c) Write the python code to open and read the file "***sequence.seq***" which is located in ***C:\files*** path on a Windows operating system. Search for those lines which start with C and print them out. Do not forget to eliminate extra next line at the end of each line before printing. [3 marks]

Continued...

QUESTION 3 [10 marks]

a) What is difference between Python and Biopython? [2 marks]

b) What are the different types of IUPAC Alphabet attributes in Biopython?

[4 marks]

c) Write a python script to do complements and reverse complements for the given DNA sequence using Biopython module?

"AGTAACCCTTAGCACTGGT" [4 marks]

Continued...

QUESTION 4 [10 marks]

a) Given x<-matrix(c(3,4,-2,6),nrow=2,byrow=T)
and y<-matrix(c(1,3,0,9,5,-1),nrow=3,byrow=T)
write the x and y matrices, then calculate y * x and y%*%x [4 marks]
>x (output?)
>y (output?)
>x*y (output?)
>y*x (output?)
> y%*%x(output?)

b) Write R code to draw a histogram of table D$wg. Add a title "Weight gain" to the table. Label the X axis as "Gain", and the Y axis as "Freq". Add "Red" color to the histogram. [4 marks]

c) Explain briefly what each of the following R functions does. [2 marks]
length(object)
class(object)
c(object, object, ...)
fix(object)

Continued...

QUESTION 5 [10 marks]

a) Write an R program to read a sequence from the user. The program counts the number of bases, "A","T","C", and prints the length of the sequence, bases, and the number of occurrences. [4 marks]

b) Translate a DNA sequence from a fasta file. Calculate its GC content as well. Save the translated sequence(AA sequence) as fasta file. [4 marks]

c) What commands can you use to get **help** in **R** programming. (State at least two) [2 marks]

End of Paper.